1953

Planning a testing program for the elementary schools of Forsyth Montana

Emile Louis Perey
The University of Montana

PLANNING A TESTING PROGRAM FOR THE ELEMENTARY SCHOOLS OF FORSYTH, MONTANA

by

EMILE L. PEREY

B. A. Montana State University, 1931

Presented in partial fulfillment of the requirements for the degree of Master of Education

MONTANA STATE UNIVERSITY

1953

Approved by:

Chairman, Board of Examiners

Dean, Graduate School

June 10 1953

Date

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION AND DELIMITATION OF THE FIELD OF STUDY

The assumption may fairly be made that the schools of Forsyth, Montana, are typical of schools in the small towns of the state. At any rate, certainly, the similarities outweigh the differences. Forsyth is a community of approximately 2,000 people. The principal industries supporting the town are irrigated and dry land farming, cattle and sheep raising, railroading, and the usual retail businesses catering to the needs and tastes of those occupied in the foregoing.

The schools of the town are organized on the 8 - 4 plan. There are at present between 120 and 130 students in the high school and between 330 and 360 in the elementary. An exact figure cannot be given because there is always a small transient population; this is composed chiefly of itinerant workers in sugar beet and other agricultural crop work. The population variation can also, to some extent, be attributed to oil exploration activities being conducted at the present time although no oil has yet been found in the immediate vicinity.

This paper will deal exclusively with the testing program for the first eight grades of the school system.

In these eight grades, thirteen classroom teachers are employed as well as one music teacher who spends five sixths of her time in these grades and the remaining period in the high school. The superintendent acts as principal of the elementary school in addition to his other duties.

The above gives a short, though accurate, description of the size of the educational system in Forsyth. Other parts of the picture such as teacher training and experience which must be considered in discussing a testing program will be mentioned later.

A discussion of a testing program for any school must necessarily lead to consideration of other parts of the educational whole. Philosophy of education, objectives of education, teacher training, educational leadership, public relations, and, to some extent, finance, all come into the pattern. An attempt will be made to minimize discussion of the above in this paper and to keep to the subject of tests and testing procedures. Obviously, though, a treatment of philosophy or policy and objectives is necessary because, unless a system can define its objectives, the results of its educational efforts cannot be measured. School philosophy and its concomitants will be discussed in chapter two.

The original thought was to limit consideration of tests to those practical in a small school. However, the

thought presents itself that any test useful in a large school may also be useful in a small school. The practicality of certain tests may depend on the availability of teachers who are trained in the administration of these tests. The Stanford-Binet Intelligence Test, for example, should be administered only by a trained person if the results are to be useful and valid.

As a point of departure for deciding what type of tests to use, Ross[1] gives the following: group intelligence tests, achievement batteries, achievement tests in special selected subjects. These may be taken as the core for the testing program, but other types of tests seem helpful. Because neither group intelligence tests nor achievement batteries are completely dependable, individual intelligence tests and supplementary forms of achievement batteries will be needed for retesting certain pupils. Too, since present-day education stresses other outcomes than factual knowledge, tests or scales of attitude, personality, and civic beliefs cannot be ruled out without earnest study. A reading readiness test should be given so the first grade teacher may know better what can be expected of each child. Lastly, in order that pupils may be helped whenever and wherever possible,

---

[1]C. C. Ross, Measurement in Today's Schools (New York: Prentice-Hall, Inc., 1947), p. 176.

diagnostic tests must be studied and their use recommended in those subject areas in which they are considered valid.

The following, then, are the types of tests that will be studied in this paper: group intelligence tests, achievement batteries, achievement tests in special subjects, individual intelligence tests, attitude scales, personality scales, civic belief scales (or similar measures of democratic indoctrination), reading readiness tests, and diagnostic tests in special subject areas.

CHAPTER II

LOCAL BACKGROUND

1. PRESENT TESTING PROGRAM

In the past, giving standard tests at the end of each semester has been customary in the Forsyth Public Schools. Until four years ago, this custom was carried on as a matter of course, unquestioned either by teachers or by a succession of administrators each carrying on a program inherited from a predecessor.

The chief apparent purpose of this testing program was only to determine whether classes and individual students were above, on, or below the norms accompanying the tests. Diagnostic possibilities were ignored except insofar as to decide that if a class were lower than usual, something should be done about it. Teachers also used the tests occasionally to show that certain students were excelling in their work or to show that certain others were not meeting the requirements of the norm. At times, test scores were used to reinforce teachers' judgments when failures among students were questioned by parents or by the administrator.

Each class in the school was always well above the stated norm. Teachers and administrators recognized, in a

general way, that the norms are based on a population with standards differing from local standards and are, therefore, not entirely applicable. Nevertheless, testing time usually resulted in some prideful comment by both faculty and patrons.

The Metropolitan Readiness Tests have usually been used for entrants into the first grade although, sometimes, teachers preferred to ignore them. Last year, for the first time, another test was ordered for use; the Science Research Associates Primary Mental Abilities Test for ages five to seven. There was no really valid reason for the change except that the SRA PMA tests are more attractive, more readable. By the same token, the choice could as easily have been the California Test of Mental Maturity, Pre-primary Series, or the Lee-Clark Reading Readiness Test.

Until last year, (1951-52), the achievement tests used in the program were the Stanford Achievement Test Batteries in grades three through six and the Metropolitan Achievement Test Batteries in grades one and two. The Stanford Batteries are in five forms: D, E, F, G, and H. These were used in rotation and, thus, no form was used in any grade oftener than once every two and one-half years.

Although a battery for grade one and first half of grade two is now printed, there were no Stanford Achievement Tests available for these levels until this year. For that

reason, another test was needed here to complete the program and the Metropolitan Achievement Test Batteries, Primary 1 and Primary 11 were used. These are produced in three forms: R, S, and T, and were used in rotation as above.

Last year, too, the California Achievement Test Batteries were substituted for the Stanford and Metropolitan Batteries in grades one through six. Again, as in the case of the readiness tests, the reason was the more attractive composition of the California series plus ease of scoring with the California Scoreze sheets. The claims made by the California Test Bureau that the tests are diagnostic in nature was also a factor in the choice.

Notice will be accorded the fact that the seventh and eighth grades are not included in the testing program. This is because these two grades are under the supervision of a teacher who has been in the system for many years and who feels that, since the tests invariably support her judgment, giving them would be a waste of time, effort, and money. Because she is an excellent teacher who watches the progress of her pupils carefully, her belief is probably correct to a high degree. Authoritative support of such an attitude is found in the following: "'. . . day to day observations by alert teachers provide the most significant evidence of pupils' understandings.'"(this is) precisely the way students

will be rated in later life . . ."[1]

In the spring of 1952, the World Book Company selected the Forsyth schools as one of the systems in the area to assist in the validation of the new form J, a revision of the Stanford Achievement test series. Thus, an additional evaluative instrument was used in the year mentioned.

Since 1947, the Montana State Department of Public Instruction has been emphasizing a guidance program; the Forsyth schools have followed the suggestions in testing. This is supplemental to the program described above and, because of duplication, is somewhat wasteful. The thought here was that the State Department will eventually institute a complete program and that, by following the recommendations from year to year, the Forsyth schools could gradually drop their old program and be in accord with the state program.

In accordance with the state, the Forsyth schools in 1948, 1949, 1950, and 1951 used the Coordinated Scales of Attainment in the sixth grade and the Haggerty Intelligence Examination in the fourth grade.

For the years mentioned above, it was the policy of the State Director of Occupational Information and Guidance to

---

[1]Warren G. Findley and Douglas E. Scates, "The Measurement of Understanding," Forty-fifth Yearbook of the National Society for the Study of Education, Part I (Chicago: The University of Chicago Press, 1946), p. 45.

recommend specific tests. More recently, the policy was to recommend several tests and to mention several others as acceptable, thus leaving the final choice to the different schools. The recommendations are expanding to different grade levels and other types of tests yearly; eventually it is hoped that the recommendations will encompass a complete program. At present, an achievement battery is recommended for the sixth grade, a reading test for the fifth grade, and an intelligence test for the fourth grade. For the years 1951-52 and 1952-53, the Forsyth schools accordingly administered the California Achievement Test Battery in the sixth grade (this fits in with the program described above), and the California Reading Test in the fifth grade, and the California Test of Mental Maturity in the fourth grade.

II. TEACHING STAFF: SOME CURRICULAR IMPLICATIONS

A testing program is only a part of an educational system and cannot be divorced from the other parts. As Jordan says, "The process of education includes three main divisions: (1) the determination of goals or objectives, (2) the manipulation of materials and methods so that these objectives are achieved, and (3) the evaluation or appraisal of results obtained."[2] Unquestionably, the teachers who do

---

[2]A. M. Jordan, Measurement in Education: An Introduction (New York: McGraw-Hill Book Company, 1953), p. 3.

the "day to day" manipulation of materials and methods are the most important part of any educational system.

No matter what type of curriculum may be decided upon, no matter what the stated school philosophy may be, no matter what outcomes may be considered desirable by parents, community, school board, state department, or administration, the final implementation of these falls upon the teacher. J. Paul Leonard, in describing the meditations of the Committee of the Function of Science in General Education, part of the Commission of Secondary School Curriculum of the progressive Education Association says;

> The commission while freeing itself from the restraints of college admissions, found itself entangled in another restricting web as confusing as the one from which it had loosed itself. This web was the subject boundaries themselves to which teachers tenaciously clung. Instead of organizing a committee to prepare a unity of these various fields, each man spoke for his own field and tried to examine the materials and methods used to see how by still teaching the field he could contribute to the general education of youth. The trap of an apologia for each subject was ever present and it snapped many times during the preparation of the various reports.[3]

Although the above was written about a high school curriculum, it is equally applicable to an elementary study. Thus, it becomes apparent that without teacher indoctrination

---

[3]J. Paul Leonard, Developing the Secondary School Curriculum (New York: Rinehart and Company, 1946), p. 185.

and cooperation any curriculum change will be ineffectual. A study of the teachers involved becomes necessary because teacher attitude will vitally affect any attempted change in the educational program.

The thirteen teachers who will be in charge of the elementary rooms next year in Forsyth have educational backgrounds as follows:

	Training	Experience (years)
Teacher A	3 years	14
Teacher B	2 years, 2 quarters	6
Teacher C	2 years, 1 quarter	20
Teacher D	2 years	14
Teacher E	2 years, 2 quarters	15
Teacher F	2 years, 2 quarters	21
Teacher G	2 years, 1 quarter	4
Teacher H	3 years	10
Teacher I	4 years	3
Teacher J	2 years, 1 quarter	2
Teacher K	2 years, 1 quarter	40
Teacher L	4 years	14
Teacher M	2 years, 1 quarter	12

The average time spent by this group in training for the teaching profession is slightly more than two years and two quarters; the average number of years experience is slightly more than thirteen.

Necessarily, most of this time in training must have been spent in the "working" curriculum of teacher training institutions: methods and subject matter. Further, in several cases the training was completed a number of years ago; thus, it is probable that any child psychology, curriculum, or philosophy courses, "theory" courses they took are somewhat

outdated at the present time. Recognition must be given, however, to the fact that many of the teachers faithfully attend summer schools each year for short workshop courses or, occasionally, for complete sessions.

Nevertheless, it is probable that many of the teachers have given little thought to the newer doctrines in education and that they know little about any but the subject-centered curriculum. As Kandel says, "No theory of education, however sound, can be better than the teachers who are appointed to put it into practice."[4] Caswell, too, in discussing the role of the administration and that of teachers in affecting curriculum change, says:

> Preparation of written statements is only a small aspect of the problem. The organization of the curriculum rests fundamentally on the understandings and beliefs of educational workers and parents. It matters little what is stated to be the basis of curriculum organization if teachers and parents believe otherwise. The soundness of the point will be recognized in examining operational school situations. If a teacher believes in a subject matter curriculum, he will inevitably tend to develop that type of curriculum regardless of all else. The superintendent or the principal may demand that all teachers follow the activity concept; the daily schedule may be organized in large blocks of time; but the actual experiences of children will continue to be largely subject centered. A teacher cannot submerge his fundamental beliefs and convictions under any general plan of curriculum organization. Thus the point of view of teachers is basic, and procedures designed to foster the development of improved plans of curriculum organization

---

[4]I. L. Kandel, The Cult of Uncertainty (New York: The Macmillan Company, 1943), p. 95.

> must be based for success on the development of improved understanding and insight on the part of the educational staff.[5]

Since the teachers' background does not include enough training in the more recent theory, it is apparent that they must receive inspiration, indoctrination, and assistance if any progressive change is to be made. The initiative for changing school philosophy to encompass child-centered rather than subject-centered teaching, teaching for life rather than teaching for the purpose of passing subjects and surpassing standard test norms, must come from some leader. If for no reason other than that a better educational leader is lacking in the community being considered, the administrator will have to accept responsibility for a study of curriculum building, child psychology, and testing. His responsibility will be to arouse interest among the teachers, to combat apathy, and to enlist the aid of those who can assist with the program. An administrator must be prepared to accept the responsibility of leadership. Quoting Aiken, "Effective democratic leadership is essential. The principal is the one who would be expected to lead. That school is fortunate whose principal has the capacity and skill to be the educational leader."[6]

---

[5]Hollis L. Caswell and A. Wellesley Foshay, Education in the Elementary School (New York: The American Book Company, 1950), p. 254.

[6]Wilford M. Aiken, The Story of the Eight-Year Study (New York: Harper and Brothers, 1942), p. 134.

Another reason why this must be the job of the administrator is that the teaching staff is continually changing. Year to year continuity is more likely if the administrator is the motivating force; in fact, continuity is impossible otherwise. If the administrator in this case does not initiate change, the alternative is stagnation.

> Nor can teachers avoid responsibility. When the Eight-Year study was started, some schools selected a few members of the faculty for the new work. The others, who were not consulted felt left out. This resulted in division and misunderstanding. In some cases it led to jealousy, bitterness, and a sabotage of the new work.[7]

Help in an in-service training program such as that suggested above can be obtained from professors of Eastern Montana College of Education. The college is always glad to assist in programs of this kind in schools of this area.

III. IMPORTANCE OF THE PROBLEM

The previous rather lengthy discussion may seem far removed from testing, the title of the paper, but such is not the case. That the testing program of a school must coincide with the educational philosophy of that school is axiomatic, and is clearly indicated both by common sense and by literature in the field. A quotation from Saucier will illustrate:

---

[7] Ibid., p. 128.

> In the past, too often writers on measurement have attempted to judge the worth of an examination apart from the objectives of education and the teaching procedure. They have written extensively about validity, reliability, objectivity, economy of time and the like, with little or no regard for what the school was supposed to be doing for the child. Statistical procedure does not suffice for the appraisal of any means of evaluation if it is appraised in isolation.[8]

Lee and Lee, too, say, "The main concept guiding evaluation should be in terms of the extent to which pupils have attained the objectives of elementary education."[9]

A testing program has no purpose if the school has no defined philosophy, no defined objectives. A school in which the student is recognized as an individual, in which the readiness principle operates cannot depend entirely upon factual tests to determine whether or not its aims are being achieved because such a school is not exclusively interested in teaching facts. Attitudes, habits, personal and social adjustment, study skills as well as factual knowledge are prominent in today's educational philosophy. There must, then, be other means of evaluation instead of or in addition to factual examination. A quotation from Webb and Shotwell will serve to illustrate the above point of view.

---

[8]W. A. Saucier, Theory and Practice in the Elementary School (New York: The Macmillan Company, 1951), p. 426.

[9]J. M. Lee and Doris May Lee, The Child and His Curriculum (New York: Appleton-Century-Crofts, 1940), p. 596.

> After the goals are adequately worked out in both general statements and specific details comes the selection of methods of evaluation by which the teacher can determine how effective the desired outcomes are being realized.[10]

Also, quoting from a bulletin of the California State Department of Education:

> The kinds of objective govern the selection of appraisal techniques.[11]

The traditional school, placing chief emphasis on the teaching of facts, can, more logically, use achievement tests. Even such schools, however, should be careful to interpret achievement test results on a local basis rather than to compare local norms indiscriminately with national norms and to draw conclusions from the comparison. In this regard, the following seems pertinent:

> The fact that a class scores above or below the national norm may or may not necessarily indicate superior or inferior school work . . . Test situations must be interpreted in the light of local school conditions.[12]

A statement from Scates seems applicable to a discussion of the philosophy of testing in education.

> The time has come when we should cease to be primarily interested in comparing one child with another, one class with another, or any class with a norm. We should be primarily interested in comparing each child with himself,

---

[10] L. W. Webb and Anna Markt Shotwell, Testing in the Elementary School (New York: Farrar and Rinehart, Inc., 1939), p. 15.

[11] "Evaluating Pupil Progress," Bulletin of the California State Department of Education, Vol. XXI, No. 6, April, 1952, p. 5.

[12] Ibid., p. 31.

> with his past record, and with his potentialities. To center attention elsewhere is to miss the point--to miss the service that tests can render.[13]

Another trap into which a system might fall through use of standard tests is that of evaluating teachers by use of the norms of their classes. This was standard practice in Forsyth not long ago; it is a flagrant misuse of tests. Teachers, in fear of losing their jobs often resorted to drill from old tests. The result of this sort of teaching can only be rote memory of isolated facts. Present research indicates that drill, as such, probably does not achieve even the limited objective of retention until examination time.[14] Certainly it is not conducive to an enjoyable learning situation nor does it promote such outcomes as intellectual curiosity and logical reasoning.

Thus, the need for a school philosophy is clear and, in line with this need, it becomes apparent that tests cannot be chosen on a hit or miss basis if they are to be of any value; they must be made the object of careful study. A testing schedule must be worked out in order that waste of time and effort can be avoided. Teacher indoctrination in

---

[13]Douglas E. Scates, "The Improvement of Classroom Testing," Review of Educational Research, 8:532, January, 1939.

[14]Saucier, op. cit., p. 102.

the proper use of tests must come through in-service training and democratic discussion of school philosophy and objectives. Installation of a testing program should bring with it a renewed interest in school matters for all concerned. Considering other costs, the monetary expenditure would not be great. The cost in teacher time is another matter but if teachers can be brought to see the purpose of in-service training, the time spent will pay great dividends for the school system.

CHAPTER III

I. THE PROBLEM

As has been mentioned above, separation of a testing program from the objectives of education is pointless. Similarly, objectives of an educational program, to be of value, must be translated into a workable curriculum. The objectives are the desired outcomes; the curriculum is the means used to attain these outcomes. The term curriculum involves not only a list of subjects, in turn implying a list of objectives for each subject, but also the methods used in treating these subjects. Alberty says, "The activities that are provided for students by the school constitute its curriculum. It is by means of these activities that the school hopes to bring about changes in the behavior of students."[1]

Therefore, statements of objectives and planning for what the curriculum is to include must precede formulation of a testing program. As has been pointed out in Chapter II, the entire staff must be active in the process of formulating a new or changing an old curriculum. The actual choice of the instruments to be used in evaluating the curriculum comes after a thorough and mutual consideration of objectives

---

[1]Harold Alberty, Reorganizing the High School Curriculum (New York: The Macmillan Company, 1950), p. 95.

and methods. The chief value of considering a testing program may be the resulting renewed interest and vigor of the faculty. The process of deciding upon a series of tests may be of more value to the school than the tests themselves.

The possibility exists that the whole community may and, perhaps, should be drawn into the process. Complete books discussing community participation in curriculum change are available.[2] Discussion of such community organization is, however, beyond the scope of this paper. Should such an eventuality develop, it would be welcomed as a change for the better from the apathy toward school matters that characterizes the community concerned at present.

At this point, several authoritative quotations concerning criteria for the selection and use of tests will not be amiss. Lincoln and Workman suggest the following: the test should be reliable, valid, well standardized, objective, easy to administer, and economical in both time and money.[3]

Quoting from Ross, "A good testing program should be supplementary not duplicative, usable not confusing, economical not burdensome, comprehensive not sporadic, suggestive not

---

[2]For example see Alice Miel, Changing the Curriculum: A Social Process (New York: D. Appleton-Century Company, 1946)

[3]E. A. Lincoln and L. L. Workman, Testing and the Use of Test Results (New York: The Macmillan Company, 1935), pp. 39-44.

dogmatic, progressive not static."[4]

Traxler gives fifteen criteria of a good testing program. A good testing program must be comprehensive, inclusive of all students, tests must be given at regular intervals, tests must be well-timed (for example, diagnostic tests must be given early enough so corrective measures can be taken after results are known), tests should be comparable, tests must be consistent with the school's objectives and curriculum, tests must be carefully administered (this involves the training of teachers), tests must be carefully and accurately scored, the results must be interpreted in terms of appropriate norms, tests must be reported to teachers in understandable terms, records must be kept on individual cumulative records, test scores must be related to other information, there must be special testing as needed for individual cases, there must be an in-service-training program for educating teachers in the proper use of test results.[5]

At this point, the problem should be clear. The rest of this chapter will be devoted to a discussion of the actual tests to be considered. These will be taken up in the following order: group intelligence tests, individual intelligence tests, reading readiness tests, achievement batteries, achievement and diagnostic tests in specific

---

[4]C. C. Ross, Measurement in Today's Schools (New York: Prentice-Hall, Inc., 1947), p. 176.

[5]Arthur E. Traxler, "Fifteen Criteria of a Testing Program," Clearing House, 25:3-7, September, 1950.

areas, and personality scales.

Consideration of age levels appropriate for the various tests, frequency of the use of the various tests, most appropriate time of the year, week, and day for testing, and recording and use of the test results is also necessary.

II. THE STUDY

Group intelligence tests. That intelligence tests are of primary importance is illustrated by the following quotation from one of Traxler's works: "If it were necessary to limit the testing program to a single type of test probably a mental ability test would be selected, preferably one designed to give a diagnostic picture of capacity."[6]

An examination of six educational publications[7] devoted wholly or in large part to the subject of testing

---

[6]Arthur E. Traxler, et. al., Introduction to Testing and the Use of Test Results in Public Schools (New York: Harper and Brothers, 1953), p. 16.

[7]The publications referred to are:

Woodrow A. Brown, "Testing in Pennsylvania's Public Kindergartens." Test Service Bulletin No. 67, New York: The World Book Company, undated;

Harry A. Greene, Albert N. Jorgensen, and J. Raymond Gerberich, Measurement and Evaluation in the Elementary School (New York: Longmans Green and Company, 1942), pp. 210-211;

J. M. Lee and Doris May Lee, The Child and His Curriculum (New York: Appleton-Century-Crofts, Incorporated, 1940), p. 661;

W. A. Saucier, Theory and Practice in the Elementary School (New York: The Macmillan Company, 1951), p. 199;

George Spache, "Using Tests in the Small School System," Educational and Psychological Measurements, 1:99-109, May, 1946;

L. W. Webb and Anna Markt Shotwell, Testing in the Elementary School (New York: Farrar and Rinehart, Incorporated, 1939), pp. 99-108.

shows that certain tests consistently received recommendation. The next paragraph gives the ranges of the various tests and the frequency with which they were recommended.

The Pintner-Cunningham Primary Mental Test with a range of kindergarten through grade two and the Detroit Beginning First-Grade Intelligence Test used only for entrants into the first grade were each recommended in three places. The Pintner-Durost Elementary Test with a range of grades two and three, the California Test of Mental Maturity with six forms ranging from kindergarten through adult, and the Kuhlman-Anderson Intelligence Test with a range of beginning first-grade through grade twelve were each recommended in two places. The California Short-Form Test of Mental Maturity with two forms ranging from kindergarten through grade three, the Detroit Primary Intelligence Test with a range of grades two through four, the Detroit Advanced First-Grade Intelligence Test with a range of late grade one and early grade two, and the Pintner Intelligence Test with a range of grades four through eight were each recommended in one place.

Other comments regarding this phase of testing must be mentioned here. In writing about the use of tests at the primary level, Saucier follows his lists of tests with,

> To discover the maturity or development of first grade pupils for reading, authorities advocate giving them an intelligence test, a readiness test and a list of questions of their interests. However, nothing can take the place of the teacher's judgment in the matter.[8]

---

[8]Saucier, op. cit., p. 204.

Emphasizing the importance of delaying reading instruction for those who are not ready, he says,

> Reports indicate that the unreadiness of the children in this (first) grade is between 25 and 50 per cent. This means that, in a typical first grade, a teacher of thirty pupils is justified in expecting about ten of them to be unready for reading the whole school year.
>
> What we need to realize is that the teacher of reading by not reckoning with the problem of reading readiness creates the problem of the non-reader or the retarded reader.[9]

Emphasis is given this problem of reading readiness and its determination through tests because inability to read is of the greatest importance; indeed, this ability is requisite to success in the modern school program. No figures are available concerning the number of pupils who are retarded throughout their school careers, often becoming serious discipline problems, merely because of their inability to read. If preventive measures can be promoted through proper use of tests in the first grade, no better result could be asked of the entire program.

Spache,[10] a New York State elementary principal, states that he is satisfied with none of the group intelligence tests at the elementary level. For this reason, he uses the Porteus Maze and the Stanford-Binet Intelligence Test, both individually administered.

Those tests most frequently mentioned in the six publications listed above are, again, the Pintner-Cunningham

---

[9] Ibid., p. 203.

[10] George Spache, op. cit., p. 104.

Primary Test and the Detroit Beginning First-Grade Intelligence Test. The Pennsylvania study showed great preference for the Pintner-Cunningham Primary Test. The only tests mentioned that have forms ranging from kindergarten through grade eight, the grades concerned in the problem, are the Kuhlman-Anderson Intelligence Test and the California Tests of Mental Maturity, each mentioned in two places.

Using the above information by itself makes the choice difficult. The factors of usability, ease of administration, ease of scoring, economy of time and money, validity, reliability, and attractiveness of format must be considered. The California Tests of Mental Maturity seem to fit the above. In regard to these tests, Kuhlman says,

> . . . we believe the unabreviated batteries are to be classed as among the very best on the market for determining general levels of mental maturity. It is gratifying to see authors with the courage to offer tests that take more than a single class period to give and who do not attempt to get the maximum economy in time and dollars, by sacrificing everything necessary to attain this end.[11]

The desirability of having a test for use at several age levels that would offer comparable scores rather than using completely different tests, the scores of which are not comparable, seems worth considering. Further, the teachers concerned are already somewhat familiar with this test, the new Scoreze pads used above the primary level with the California Tests eliminate error in scoring, and it is

---

[11]Oscar Krisen Buros, editor, The Mental Measurements Yearbook, Article concerning The California Test of Mental Maturity by F. Kuhlman, (Highland Park, New Jersey: The Mental Measurements Yearbook, 1941), p. 209.

already part of the recommended state program with which the school is planning to comply.

Individual intelligence tests. Examination of four tests[12] written exclusively on testing reveals a number of individual intelligence tests. Those that were mentioned but have been superseded by newer editions and those that pertain to age levels not concerned in this study will be ignored here. There seems little doubt that the best-known and most widely used test is the New Stanford-Binet Scale, completed in 1937.

The tests that will be mentioned are all usable at the age level with which we are concerned. The New Revised Stanford-Binet Intelligence Test is mentioned in all four texts, the Kuhlman Revision of the Binet Scale is mentioned in two texts, and the Porteus Maze Test in one text, and the Wechsler-Bellevue Intelligence Scale in one text.

The Pintner-Patterson Scale of Performance Tests are mentioned in three texts, and the Arthur Point Scale of Performance is mentioned in two texts. The latter two are performance scales intended to eliminate the verbal factor and to supplement the Stanford-Binet.

Greene says, "The New Revised Stanford-Binet Tests of Intelligence is today, as was its 1916 predecessor, the

---

[12]The four texts referred to are: Greene, Jorgensen, and Gerberich, op. cit., pp. 206-209; A. M. Jordan, Measurement in Education: An Introduction (New York: McGraw-Hill Book Company, 1953), p. 335-371; Ross, op. cit., pp. 36-41; Webb and Shotwell, op. cit., pp. 74-89.

best known and most widely used individual test of general intelligence in America."[13] Jordan, too, recommends it highly. He says, "Most workers believe (the Terman-Merrill Revision of the Stanford-Binet Intelligence Test) the best test of its kind ever constructed."[14]

In the light of the above, there can be no doubt that the New Stanford-Binet Intelligence Test must be the choice.

Reading readiness tests. This type of test is taken up at this point in the paper because of its close connection with intelligence tests at the primary level. Jordan believes that, "While intelligence tests are of value in beginning formal instruction and number work, they do not predict final achievement marks as well as tests of reading readiness."[15] Saucier[16] believes both group intelligence tests and reading readiness tests should be used. Other authors do not concur with Saucier. Group intelligence tests do not appear to be valid enough at this level to justify their use.[17]

Examination of five texts[18] shows the most often-

---

[13]Greene, Jorgensen, and Gerberich, op. cit., p. 206.

[14]Jordan, op. cit., p. 361.

[15]Jordan, Ibid., p. 97.

[16]Cf. ante., p. 21.

[17]Spache, op. cit., p. 194.

[18]The five texts referred to are: Greene, Jorgensen, and Gerberich, op. cit., pp. 333-336; Gertrude Hildreth, Learning the Three R's (Minneapolis: Educational Publishers, Incorporated, 1947), pp. 185-186; Jordan, op. cit., pp. 97-102; Lee and Lee, op. cit., p. 660; Webb and Shotwell, op. cit., pp. 132-134.

mentioned reading readiness tests to be as follows: the Lee-Clark Reading Readiness Tests, the Monroe Reading Aptitude Tests, and the Metropolitan Readiness Tests were each mentioned in four places; the Gates Readiness Test was mentioned in three places; the Betts Ready to Read Test and the Stone-Garver Classification Tests for Beginners in Reading were each mentioned in two places; the Durrell-Sullivan Reading Capacity Test was mentioned in one place. There was general agreement that the Monroe Reading Aptitude Test was the best for group testing, though a part of it must be administered individually. At the grade level involved, however, individual testing is almost a necessity in any case. Even in group testing, the teacher must pay extreme attention to individuals and large classes cannot be tested together. The Van Wagenen Reading Readiness Tests are deemed the best of individual tests to be used.

In connection with the Monroe Reading Aptitude Tests, one question arises. On the motor phase of the test, the child is asked to draw a pencil line through a series of dots and dashes; the score on this part of the test is determined by how many of the dots are completely covered by the pencil line. Missing several of the dots or failing to cover them completely would seriously affect the child's score. Here, it seems that the score might depend upon how dull the child's pencil is when he takes this part of the test.

Achievement test batteries. In general, there are two types of achievement test batteries. One type covers the

various subject areas taught in schools; while the other type concentrates on what might be called the fundamentals which must be learned, whatever the school's philosophy may be.

The first type is exemplified by the Stanford Achievement Test Batteries and the Metropolitan Achievement Test Batteries. Since the publication of the new Stanford Achievement Test Battery, Form J, each of the above attempts to cover the entire elementary field, grades one through eight.

The second type of achievement test is exemplified by the Iowa Every-pupil Tests of Basic Skills and the California Achievement Tests.

Regarding the Stanford Achievement Tests, Jordan says, "Since the publication of these tests, or this set of tests, many test batteries have been constructed, but all of them have certain characteristics in common with the first achievement test battery."[19] The Metropolitan Achievement Tests and the Stanford Achievement Tests are alike in many respects.[20] Each has sections devoted to the subject matter taught in most schools--arithmetic, history or social sciences, punctuation, and so on. Profiles can be drawn from the results of each test, and grade and age placement can be read from the tables.

The California Achievement Test has one apparent weakness that stems from the fact that the single primary battery is used for grades one, two, and three. A test

---

[19]Jordan, op. cit., p. 80.

[20]Ibid., p. 82.

that is appropriate for grade three cannot possible be be appropriate for grade one. Even if the items are arranged in order of increasing difficulty, a sensitive first-grade pupil would inevitably feel frustrated upon being confronted with so many impossibly difficult questions. There is too great an educational difference in these grades for any one test to fit all three.

The Iowa Every-pupil Test of Basic Skills appears to be the most thorough in the field. There are two batteries, one for use in grades three through the first semester of grade five, and one for use in the second semester of grade five through grade nine. Each battery is composed of fourteen different tests which, as mentioned before, emphasize basic study skills rather than subject matter. These tests are available either as complete batteries or as separate tests.

Because of the cost and the time involved in administering, this study will recommend that the entire battery be given at three levels only; grades four, six, and eight. That general achievement tests are of value below the third grade is questionable,[21] and none will be given in the system in question except at the specific request of the teacher concerned. In grades three, five, and seven, the use of the Stanford batteries will be presented to the teachers as advisable. After all is said about testing in the modern program, one cannot but believe that there is

---

[21]Ross, op. cit., p. 176.

some student motivation in being faced with a standard test. One must remember, though, that the results should be discussed with the student so he can know his own weaknesses, and that the purpose of standard tests is not to decide the issue of promotion or retention.

Achievement and diagnostic tests in specific subjects. These are a necessary corollary to the testing program because, for various reasons, it may be desirable to test certain students a second time in certain subjects. For example, if a student's score on a phase of the regularly scheduled battery varies too widely from the teacher's personal evaluation of his ability, he should be re-tested in that phase.[22] There may be many reasons other than lack of ability or knowledge for the student's low score and he should have a chance to try again. One must be careful not to form opinions on the basis of a single test. For this purpose, assuming that the Iowa Tests are used for the original program, a separate test of the California Battery in the appropriate field can be used.

Diagnostic tests are included in this section because it appears difficult to consider them apart from subject matter tests. Different writers and companies that produce tests present confusing concepts of diagnostic tests. Tests are advertised measuring achievement and, at the same time, as being diagnostic. There seems to be a fallacious concept

---

[22] Ibid., p. 186.

here. Achievement tests offer too few items in each skill to be truly diagnostic. They may be diagnostic insofar as they indicate a weakness in an area or insofar as they show a class to be below the expected achievement level. However, true diagnosis not only discovers weakness but points to the reason for it. In this connection Greene says,

> It is interesting to learn, as a result of giving tests in the classroom, that a pupil of the entire class is below standard in a subject, but unless it is learned with some exactness what causes the low level of achievement, the testing program will do little if anything more than supply interesting information.
>
> Accurate diagnosis of class and individual pupil difficulties, coupled with application of specific remedy, is the heart of enlightened use of exact methods of teaching.[23]

Lincoln and Workman differentiate between diagnostic and achievement tests in the following quotation:

> The achievement test may give five to ten examples in addition, each of them different from any of the others. If the pupil fails on these or on most of them, it is not possible to tell why he fails. . . the diagnostic test must be so constructed that the particular trouble of the pupil will be apparent; therefore, it must contain many more types of examples and many more examples of each type than the ordinary test does.[24]

The Compass Diagnostic Test in Arithmetic, to cite a specific case, has twenty separate tests involving each phase of arithmetic processes; each test has from four to seven parts.

Certain subject areas do not lend themselves to this type of diagnosis. A test may show that a student is low in

---

[23]Greene, Jorgensen, and Gerberich, op. cit., p. 293.

[24]Lincoln and Workman, op. cit., p. 126.

history but it could not show why--unless the difficulty were traced back to reading. Ross says, "Most of the diagnostic tests published to date are limited to the tool subjects mainly on the elementary level.[25] In this paper, therefore, discussion of diagnostic tests will be confined to arithmetic, reading, and spelling.

Arithmetic. The only diagnostic test for arithmetic recommended in each of the four texts on testing[26] is the Compass Diagnostic Test in Arithmetic. This is really a series composed of twenty tests covering all fundamental processes in arithmetic. The easier ones can be used in grades two through eight and they grow progressively harder, the last ones being for use only in grades seven and eight. Each may be used separately to test specific skills. The test may be administered to groups. The Compass Test described above requires careful study by the teacher using it and, thus, will have to be a subject of discussion at faculty meetings.

Hildreth[27] goes into detail illustrating the complexity of complete diagnosis; this involves testing for mental ability, reading, home background, health, interests, and personality traits. Arithmetic, too, may be a reading

---

[25]Ross, op. cit., p. 370.

[26]The four texts referred to are: Greene, Jorgensen, and Gerberich, op. cit., p. 315; Hildreth, op. cit., p. 858; Jordan, op. cit., p. 229; Webb and Shotwell, op. cit., p. 184.

[27]Hildreth, op. cit., pp. 858-870.

difficulty; the passage referred to shows that, though a teacher may give the test, no teacher, alone, can diagnose thoroughly. All the texts studied were careful to bring out the point that diagnosis by itself is pointless, and all texts suggested methods of remedial follow-up.

Reading. The same four texts[28] mentioned above under arithmetic are unanimous in choosing Gray's Standarized Oral Reading Check Tests and Gates' Silent Reading Tests. The first is an individual test and comes in four forms covering grades one through eight. The second, a group test, has one series for grades one and two and another series for grades three through eight. Each series has two forms.

The California Reading Test used by the school as part of the state program was mentioned only by Jordan who commented favorably on its attempt to build both achievement and diagnostic functions into one test. Though it is not truly diagnostic, its use will be continued in the fifth grade and it will be recommended for the seventh grade. As supplemental to the California Reading Test, the Gates and the Gray tests mentioned above will be suggested for use in special cases as needed.

That reading is an area too large to be completely diagnosed by the classroom teacher is indicated by

---

[28]The four texts referred to are: Greene, Jorgensen, and Gerberich, op. cit., pp. 336-347; Hildreth, op. cit., pp. 386-387; Jordan, op. cit., pp. 108-109; Webb and Shotwell, op. cit., pp. 134-142.

Monroe[29] who, in describing a complete diagnostic examination, suggests the following areas of possible difficulty: visual difficulties, auditory difficulties, motor difficulties, vocabulary difficulties, difficulties due to faulty teaching methods in earlier grades, enviromental difficulties, and emotional difficulites. Use of the Telebinocular and Ophthalm-O-Graph is also suggested. In a school that has neither specially trained staff members nor equipment of this kind, a teacher can only do her best by choosing tests carefully and being kind, sympathetic, and tireless in her efforts to remedy suspected reading difficulties.

A discussion of reading would not be complete without reference to some of the texts on the subject. On the list is Dolch's Manual for Remedial Reading, which contains many helpful informal tests for reading difficulties together with suggestions for remediation. As an example, he gives a list of two hundred twenty words, with suggestions for teaching them, that form at least fifty-three per cent of the words used in all subjects, grades one through eight.

The following books are excellent for a study of reading:

Dolch, Edward W., Manual for Remedial Reading. Champaign, Illinois: The Garrard Press, 1946. 460 pp.

_____, Problems in Reading. Champaign, Illinois: The Garrard Press, 1949. 373 pp.

---

[29]Marion Monroe, Children Who Cannot Read (Chicago: The University of Chicago Press, 1932), pp. 79-110.

Durrell, Donald D., Improvement of Basic Reading Abilities. New York: The World Book Company, 1940. 407 pp.

Fernald, Grace M., Remedial Techniques in Basic School Subjects. New York: McGraw-Hill Book Company, 1943. 349 pp.

Gates, A. I., The Improvement of Reading: A Program of Diagnostic and Remedial Techniques. New York: The Macmillan Company, 1947. 657 pp.

Gray, William S., On Their Own in Reading. New York: Scott, Foresman and Company, 1948. 268 pp.

Monroe, Marion, Children Who Cannot Read. Chicago: The University of Chicago Press, 1932. 205 pp.

Osborn, W. J., Phonetic Elements: A Phonetic Inventory For Use in Spelling and Reading. Unpublished Pamphlet, Associated Students University of Washington Book Store, Seattle, Washington. 14 pp.

_____, Growing into Reading. New York: Scott, Foresman and Company, 1951, 274 pp.

Spelling. The same texts[30] were used for studying spelling diagnosis as were used for the study of arithmetic and reading diagnosis. In addition, Fernald's book[31] was used. Apparently difficulty in reading and difficulty in spelling are concomitant to a great extent. Those children who have difficulty in spelling often, though not always, have difficulty in reading. Hildreth does not mention specific spelling tests but devotes four chapters to the problem of proper teaching, diagnosing, and remedial teaching of spelling.

---

[30]The texts referred to are: Greene, Jorgensen, and Gerberich, op. cit., pp. 376-384; Hildreth, op. cit., pp 470-581; Jordan, op. cit., pp. 123-127; Webb and Shotwell, op. cit., p. 220.

[31]Grace M. Fernald, Remedial Techniques in Basic School Subjects (New York: McGraw-Hill Book Company, 1943), pp. 181-210.

Greene, Hildreth, Jordan, and Fernald all emphasize the teaching of spelling incidentally along with the teaching of other subjects. They suggest that spelling lessons using words out of context, words that have no connection with what the child is experiencing, reading, and writing in other subjects is not only pointless to the child but often harmful. Repetition of formal drill as such may cause the child, particularly in the primary grades, to dislike spelling. This dislike results in cumulative difficulty with the subject and may produce a non-speller or a problem speller.

Fernald points out that some children cannot visualize well and, thus, cannot profit by the usual methods of teaching. She describe simple methods of kinesthetic and auditory teaching that are often successful in these cases.

These authors insist that formal drill is not good teaching nor is forcing a child to write and re-write words. This is especially true when the child copies his own words because this often results in practice of errors. Cumulative personal lists of words which the child cannot spell easily is recommended, as is the use of tests which encourage the child to compete with himself and, thus, improve his spelling.

Neither Hildreth nor Fernald mention specific tests, but Hildreth devotes four chapters and Fernald one to proper teaching, diagnosis, and remedial teaching of spelling. Fernald's chapter is especially good.

Greene, Jordan, and Webb each mention the Buckingham Extension of the Ayres Spelling Scale, the Iowa Spelling Scale, and the Morrison-McCall Spelling Scale. The longest list of the three is the Iowa Spelling Scale, while the Morrison-McCall Spelling Scale is composed of a sampling of words from the Ayres list; eight tests of equal difficulty, each consisting of fifty words. These latter two will be recommended to the faculty for use; the Morrison-McCall Scale for frequent use in grades two through six, and the Iowa Scale for practice in the seventh and eighth grades.

Personality tests. Tests of intangibles such as personality, attitudes, beliefs, mental health, emotional adjustment, and occupational preference seem to be inextricably intermixed. Study of authoritative texts and articles on the subject of personality tests leaves one somewhat confused, since one authority may be unmercifully critical of a certain test while another praises it. This paper will be limited, in this area, to a discussion of personality tests.

Personality is measured by several types of approaches.[32] Among these are free association, observation of behavior, rating scales, and personal reports. One example of free association is the study of the responses of a subject to spoken words. Another is the study of a subject's description of what he sees in an ink blot.

---

[32]Greene, Jorgensen, and Gerberich, op. cit., pp. 247-251.

These require the services of a psychologist. Observation of behavior is, of course, used by every classroom teacher; it is also used by placing subjects in definite situations and having them watched by trained observers, who keep an anecdotal record. Much can be learned about a child by this method. Rating scales are not easy to use because their accuracy depends on the one using the scale; one observer may be lenient while another may be over-critical. Personal reports make use of questionnaires, inventories, and scales which are filled out by the subjects. This is the type that will be dealt with in this paper. There are a number of these available from the various test producing companies; some are applicable only to high school students and adults, some to upper elementary students; but perusal of testing company datalogues revealed only one, The California Test of Personality, with forms applicable to all ages from kindergarten through adult. The books used in this part of the study agreed on no test of personality. Both Lee and Lee[33] and Jordan[34] mentioned the California Test of Personality favorably. Neither Laurance F. Shaffer[35] nor

---

[33]Lee and Lee, op. cit., pp. 665-666.

[34]Jordan, op. cit., pp. 473-477.

[35]Oscar K. Buros, editor, The Third Mental Measurement Yearbook. Article concerning the California Test of Personality by Laurance F. Shaffer, (New Brunswick, New Jersey: The Rutgers University Press, 1949), pp. 55-56.

Douglas Spencer[36] feel that this test is a good one. Earl R. Gabler[37] recommended the test, but gave no good reason for doing so except the fact that it is easy to score. Pflieger[38] reports an experiment using the California Test of Personality, and feels that the results were valuable in that they furnished the teachers some understanding of children's problems.

A directed discussion of the California Test in faculty meeting may result in a reawakened awareness of the conflicts of children. That the teachers themselves take the adult form of the test will be suggested. If the awareness mentioned above is brought about, some benefit will have resulted. At the present, no personality test will be recommended for general use in the school. After more study of the subject, perhaps one will be added to the program.

Time of testing. None of the books examined suggests a special time of day for testing. However, common sense indicates that testing should take place early in the day when the pupils are rested and fresh. As to the time of year, there is general agreement that achievement tests should be given in the fall if they are to be of most value

---

[36]Ibid., Article concerning The California Test of Personality by Douglas Spencer, pp. 56-58.

[37]Earl R. Gabler, "California Test of Personality," Clearing House, 16:188, November, 1941.

[38]Elmer F. Pflieger, "Pupil Adjustment Problems and a Study of Relationships Between Scores on The California Test of Personality and the Mooney Problems Check List," Journal of Educational Research, 41:265-78, December, 1947.

to the teacher in pointing out individual pupil needs and class needs and, hence, in planning the year's work. Too, the fall testing relieves the teacher of any worry concerning her class achievement norms. If it is desired to measure progress in the grade, another form of the same test can be given toward the end of the school year.[39]

Intelligence testing should be done as early in the school career as possible, so teachers may have a knowledge of individual capacity. Ross[40] feels that is is desirable to repeat intelligence tests until at least three have been given to the student. As mentioned before, probably intelligence tests are not too valid below the third grade. Therefore, suggestion will be made that they be used in the fourth grade (complying with the state program), seventh grade, and ninth grade. Individual intelligence tests should be given as needed, whenever a teacher feels doubt concerning the accuracy of a pupil's group test score. Reading readiness tests, of course, should be given about the end of the second week of the school year when the children have become somewhat accustomed to the new situation.

No set time can be given for the various diagnostic tests; they must be available when needed. Because reading is all-important, reading tests will be given oftener than

---

[39]Reference is made to the following texts: Greene, Jorgensen, and Gerberich, op. cit., p. 107; Jordan, op. cit., p. 69; Ross, op. cit., p. 195.

[40]Ross, op. cit., p. 295.

the state program suggests. Reading tests will be suggested for the beginners in grade three, five, and seven.

CHAPTER IV

SUMMARY

REVIEW OF THE PROBLEM

The problem presented in this paper is that of planning a testing program for a small Montana elementary school. Consideration of the problem makes it apparent that a testing program has no value unless the school objectives and philosophy are considered first. This, in turn, necessitates study of the curriculum--the specific objectives of each subject and the teaching methods through which the school tries to bring about realization of the general objectives. There must be, therefore, study and discussion by the faculty resulting in general agreement of just what the school is trying to do for its pupils.

RE-STATEMENT OF THE LIMITATIONS OF THE STUDY

Obviously, the tests recommended can only be those that can be properly administered by the personnel available. Faculty discussion of the proper administration and scoring of tests and the use of test results becomes a necessity. The tests used at the inauguration of the program must be those that are somewhat familiar. This minimal program can be extended as the teachers acquire knowledge of testing.

CONCLUSIONS

The types of tests recommended for faculty study will be group intelligence tests reaching each student three times during his school career, individual intelligence tests for certain students, reading readiness tests, achievement test batteries, achievement tests for use in re-testing specific areas, diagnostic tests in arithmetic, reading, and spelling for use in special cases, and personality tests.

One conclusion resulting from this study is that, possibly, too much is expected by many educators from standard tests. Many seem to have the idea that standard tests in themselves will benefit a school in some vague way. The word standard seems to carry with it an esoteric implication; something that can be understood only by experts and is hidden from all others.

Teachers need to realize that the use of tests will do nothing for education unless results are used properly. The study of results, however, can bring about educational improvement through abandonment of poor teaching methods and by calling attention to individual pupils who may need help.

Teachers need to realize, too, that informal tests may be as good as or better than standard tests. Findley and Scates say, "In every subject matter area there are available at present many well-known procedures for the

evaluation of understanding."[1] These are: normal classroom opportunities for observation, examination of work products, written tests, pupil interviews, and observation of pupil behavior. The following seems significant enough to be quoted for the second time: ". . . day to day observations by alert teachers provide the most significant evidence of pupils' understandings."[2]

RECOMMENDATIONS

This section of the paper will deal with the specific tests that will be recommended for consideration by the faculty of the elementary school of Forsyth, Montana, together with suggestions concerning the grade levels and time of year at which the tests might best be used. A complete alphabetical list of these tests may be found in Appendix A, and another list indicating subject, type, and purpose of the tests in Appendix B.

Intelligence tests. The following intelligence tests will be recommended for use at the levels indicated:

"The California Tests of Mental Maturity," a group intelligence test, early in the school year in the fourth, seventh, and ninth grades;

"The New Stanford-Binet Intelligence Test," an

---

[1]Warren G. Findley and Douglas E. Scates, "The Measurement of Understanding," Forty-fifth Yearbook of the National Society for the Study of Education, Part 1, (Chicago: The University of Chicago Press, 1946), p. 45.

[2]Ibid., loc. cit., Italics not in the original.

individual intelligence test, for use when the teacher's judgment of the pupil varies greatly from the pupil's score on the group intelligence test.

Reading readiness tests. The following reading readiness tests will be recommended for use at the times indicated:

"The Monroe Reading Aptitude Test," a group test, for use during the third week of the first grade;

"The Van Wagenen Reading Readiness Test," an individual test, for use in re-testing first grade pupils whenever it may be thought necessary.

Achievement test batteries. These achievement test batteries will be suggested for use as indicated:

"The Iowa Every-pupil Test of Basic Skills," a group test, for use early in the school year in the fourth, sixth, and eighth grades;

"The Stanford Achievement Tests," group tests, for use early in the school year in the third, fifth, and seventh grades;

Other forms of the same tests may be used at the end of the school year in case any teacher desires to measure progress.

General achievement and diagnostic tests in specific subjects. Parts of the "California Achievement Tests" will be suggested for use in re-testing pupils in specific fields when deemed necessary.

The following tests will be suggested for use at the levels and for the purposes indicated:

"The California Reading Test," a group test, for use early in the school year in the fifth and seventh grades;

"Gray's Standardized Oral Check Tests," an individual test, and "Gates' Silent Reading Tests," a group test, will be suggested for use whenever the teacher deems it necessary;

"The Compass Diagnostic Tests" in Arithmetic, a group test that can be used in grades two through eight, will be recommended for use whenever the teacher feels it may be needed;

"The Iowa Spelling Scale" will be proposed for use in grades seven and eight and the "Morrison-McCall Spelling Scale" will be proposed for grades two through six.

Personality tests. The suggestion will be made that the teachers themselves take the adult form of "The California Test of Personality" and that they discuss the student forms with a view to possible later adoption.

General pattern of tests by grade level. The following section of the paper will present the suggested tests by grade level and time of year for testing:

Grade I.

"Monroe Reading Aptitude Test;" third week of the school year.

Grade II.

"Morrison-McCall Spelling Scale;" no definite time

of year.

Grade III.

"Morrison-McCall Spelling Scale;" no definite time of year.

"Stanford Achievement Test Battery;" early in school year; a second form at the end of the school year if the teacher desires.

Grade IV.

"Morrison-McCall Spelling Scale;" no definite time of year.

"California Test of Mental Maturity;" early in school year.

"Iowa Every-pupil Test of Basic Skills;" early in the school year; a second form at the end of the school year if the teacher desires.

Grade V.

"Morrison-McCall Spelling Scale;" no definite time of year.

"California Reading Test;" early in school year.

"Stanford Achievement Test Battery;" early in school year; a second form at the end of the year if the teacher desires.

Grade VI.

"Morrison-McCall Spelling Scale;" no definite time of year.

"Iowa Every-pupil Test of Basic Skills;" early in school year; a second form at the end of the year

if the teacher desires.

Grade VII.

"Iowa Spelling Scale;" no definite time of year.

"California Test of Mental Maturity;" early in school year.

"California Reading Test;" early in school year.

"Stanford Achievement Test Battery;" early in school year; a second form at the end of the year if the teacher desires.

Grade VIII.

"Iowa Spelling Scale;" no definite time of year.

"Iowa Every-pupil Test of Basic Skills;" early in school year; a second form at end of year if the teacher desires.

Other suggested tests not specifically scheduled but available:

"New Stanford-Binet Intelligence Tests" (all grades).

"Van Wagenen Reading Readiness Test" (first grade).

Parts of the "California Test Battery" (grades three through eight).

"Compass Diagnostic Tests in Arithmetic" (Grades two through eight).

"Gates Standardized Silent Reading Tests" (all grades).

"Gray's Oral Reading Check Tests" (all grades).

An expert in psychometry reading through this paper would feel that the treatment of the subject is cursory. He would be right. This final paragraph is being written to the accompaniment of a deep sense of inadequacy. Perhaps,

even probably, the value, if any, of this paper will come through the actual use of the ideas presented. Teachers, like most humans, tend to fall into the security of the usual, the known, the customary way. Business men, salesmen, doctors, and lawyers have conventions from which they return encouraged, filled with new ideas, and re-imbued with a renewed sense of personal worth, with the importance of their respective jobs. Teachers, too, and the term is used in a generic sense to include everyone in the educational field, need encouragement and the impact of new ideas. The hope is that this study will be one means of promoting an educational awakening among the teachers of the school concerned.

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

Aiken, Wilford M., The Story of the Eight-year Study. New York: Harper and Brothers, 1942. 157 pp.

Alberty, Harold, Reorganizing the High School Curriculum. New York: The Macmillan Company, 1950. 458 pp.

Buros, Oscar Krisen, The Mental Measurements Yearbook. Highland Park, New Jersey: The Mental Measurements Yearbook, 1941. 674 pp.

Buros, Oscar Krisen, The Third Mental Measurements Yearbook. New Brunswick, New Jersey: The Rutgers University Press, 1949, 1047 pp.

Caswell, Hollis L., and A. Wellesley Foshay, Education in the Elementary School. The American Book Company, 1950. 406 pp.

Fernald, Grace M., Remedial Techniques in Basic School Subjects. New York: McGraw-Hill Book Company, 1943. 349 pp.

Findley, Warren G., and Douglas E. Scates, "The Measurement of Understanding," Forty-fifth Yearbook of the National Society for the Study of Education, Part I Chicago: The University of Illinois Press, 1946, 338 pp.

Greene, Harry A., Albert N. Jorgensen, and J. Raymond Gerberich, Measurement and Evaluation in the Elementary School. New York: Longmans, Green and Company, 1942. 639 pp.

Hildreth, Gertrude, Learning the Three R's. Minneapolis: Educational Publishers, Incorporated, 1947. 897 pp.

Jordan, A. M., Measurement in Education. New York: McGraw-Hill Book Company, 1953. 533 pp.

Kandel, I. J., The Cult of Uncertainty. New York: The Macmillan Company, 1943. 129 pp.

Lee, J. M., and Doris May Lee, The Child and His Curriculum. New York: Appleton-Century-Crofts, Incorporated, 1950. 710 pp.

Leonard, J. Paul, Developing the Secondary School Curriculum. New York: Rinehart and Company, 1946. 560 pp.

Lincoln, E. A., and L. L. Workman, Testing and the Use of Test Results. New York: The Macmillan Company, 1935. 317 pp.

Miel, Alice, Changing the Curriculum. New York: D. Appleton-Century Company, 1946. 242 pp.

Monroe, Marion, Children Who Cannot Read. Chicago: The University of Chicago Press, 1932. 205 pp.

Ross, C. C., Measurement in Today's Schools. New York: Prentice-Hall, Incorporated, 1947. 551 pp.

Saucier, W. A., Theory and Practice in the Elementary School. New York: The Macmillan Company, 1951. 515 pp.

Traxler, Arthur E., et. al., Introduction to Testing and the Use of Test Results in Public Schools. New York: Harper and Brothers, 1953. 113 pp.

Webb, L. W., and Anna Markt Shotwell, Testing in the Elementary School. New York: Farrar and Rinehart, Incorporated, 1939. 407 pp.

B. BULLETINS

"Evaluating Pupil Progress," Bulletin of the California State Department of Education. Vol. XXI, No. 6., Henry W. Magnuson, Carl A. Larson, and Thomas A. Shellhammer. Sacramento: 1952. 184 pp.

"Testing in Pennsylvania's Public Kindergartens." Test Service Bulletin No. 67. Woodrow A. Brown. New York: The World Book Company, undated.

C. PERIODICAL ARTICLES

Gabler, Earl R., "California Test of Personality," Clearing House, XVI (November, 1941), 188 pp.

Pflieger, Elmer F., "Pupil Adjustment Problems and a Study of Relationships Between Scores on The California Test of Personality and The Mooney Problems Check List," Journal of Educational Research, XLI (December, 1947), 265-78.

Scates, Douglas E., "The Improvement of Classroom Testing," Review of Educational Research, XVIII (December, 1938), 523-36.

Spache, George, "Using Tests in a Small School System," Educational and Psychological Measurements, 1 (May, 1946), 99-109.

Traxler, Arthur E., "Fifteen Criteria of a Testing Program," Clearing House, XXV (September, 1950), 3-7.

APPENDIX

APPENDIX A

LIST OF TESTS BY TITLE

Arthur Performance Scale, (Individual intelligence test; ages five - sixteen), C. H. Stoelting Company, Chicago, Illinois.

Betts Ready to Read Battery, (Individual reading test; pre-school through college), The Psychological Corporation, New York, New York.

Buckingham Extension of the Ayres Spelling Scale, (Spelling list; grades two - nine), Public School Publishing Company, Bloomington, Illinois.

California Achievement Test Batteries, (group achievement test; grades one - fourteen), California Test Bureau, Los Angeles, California.

California Reading Tests, (Group reading tests; grades one - fourteen), California Test Bureau, Los Angeles, California.

California Short-form Test of Mental Maturity, (Group intelligence tests; kindergarten - grade three), California Test Bureau, Los Angeles, California.

California Test of Mental Maturity, (Group intelligence tests; kindergarten - adult), California Test Bureau, Los Angeles, California.

California Test of Personality, (Group personality tests; kindergarten - adult), California Test Bureau, Los Angeles, California.

Compass Diagnostic Tests in Arithmetic, (Group arithmetic tests; grades two - eight), Scott, Foresman and Company, Chicago, Illinois.

Detroit Advanced First Grade Intelligence Test, (Group intelligence tests; late first grade - early second grade), World Book Company, Yonkers-on-Hudson, New York.

Detroit Beginning First Grade Intelligence Test, (Group intelligence test; beginning first grade), World Book Company, Yonkers-on-Hudson, New York.

Detroit Primary Intelligence Test, (Group intelligence tests; grades two - four), Public School Publishing Company, Bloomington, Illinois.

Durrell-Sullivan Reading Capacity Tests, (Group reading tests; grades two - six), World Book Company, Yonkers-on-Hudson, New York.

Gates Reading Readiness Test, (Group reading readiness test; kindergarten - first grade), Teachers College Bureau of Publications, Columbia University, New York, New York.

Gates Silent Reading Tests, (Group reading tests; grades one - eight), Teachers College Bureau of Publications, Columbia University, New York, New York.

Gray's Standardized Oral Reading Check Tests, (Individual reading tests; grades one - eight), Public School Publishing Company, Bloomington, Illinois.

Iowa Every-pupil Test of Basic Skills, (Group achievement tests; grades three - nine), Houghton-Mifflin Company, Boston, Massachusetts.

Iowa Spelling Scale, (Spelling list; grades two - eight), Public School Publishing Company, Bloomington, Illinois.

Kuhlman-Anderson Intelligence Test, (Group Intelligence tests; grade one - adult), Educational Test Bureau, Minneapolis, Minnesota.

Kuhlman Revision of the Binet-Simon Scale, (Individual intelligence test; ages three months - fifteen years), Warwick and York, Incorporated, Baltimore, Maryland.

Lee-Clark Reading Readiness Test, (kindergarten - grade one), California Test Bureau, Los Angeles, California.

Metropolitan Achievement Test Batteries, (Group achievement tests; grades one - eight), World Book Company, Yonkers-on-Hudson, New York.

Metropolitan Readiness Tests, (Group readiness tests; kindergarten - grade one), World Book Company, Yonkers-on-Hudson, New York.

Monroe Reading Aptitude Tests, (Group reading test; beginning first grade), Houghton-Mifflin, Boston, Massachusetts.

Morrison-McCall Spelling Scale, (Group spelling tests; grades two - eight), World Book Company, Yonkers-on-Hudson, New York.

Pintner-Cunningham Primary Mental Test, (Group intelligence

test; kindergarten - grade two), World Book Company, Yonkers-on-Hudson, New York.

Pintner-Durost Elementary Test, (Group intelligence test; grades two -three), World Book Company, Yonkers-on-Hudson, New York.

Pintner Intelligence Test, (Group intelligence test; grades four - eight), Teachers College Bureau of Publications, Columbia University, New York, New York.

Pintner-Patterson Performance Scale, (Individual intelligence test; ages four - sixteen), C. H. Stoelting Company, Chicago, Illinois.

Porteus Maze Test, (Individual intelligence test; ages three - fourteen), C. H. Stoelting Company, Chicago, Illinois.

Stanford Achievement Test Batteries, (Group achievement tests; grades one - nine), World Book Company, Yonkers-on-Hudson, New York.

New Stanford-Binet Scale, (Individual intelligence test; ages two - adult), Houghton-Mifflin Company, Boston, Massachusetts.

Stone-Grover Classification Tests for Beginners in Reading, (Group reading test; beginning first grade), Webster Publishing Company, St. Louis, Missouri.

Wechsler-Bellevue Intelligence Scale, (Individual intelligence test; five years - adult,) The Williams and Wilkins Company, Baltimore, Maryland.

APPENDIX B

LIST OF TESTS BY TYPE

ACHIEVEMENT TEST BATTERIES

California Achievement Test Batteries, (Group; grades one - fourteen), California Test Bureau, Los Angeles, California.

Iowa Every-pupil Test of Basic Skills, (Group; grades three - nine), Houghton-Mifflin Company, Boston, Massachusetts.

Metropolitan Achievement Test Batteries, (Group; grades one - eight), World Book Company, Yonkers-on-Hudson, New York.

Stanford Achievement Test Batteries, (Group; grades one - nine), World Book Company, Yonkers-on-Hudson, New York.

ARITHMETIC TESTS

Compass Diagnostic Tests in Arithmetic, (Group; grades two - eight), Scott, Foresman and Company, Chicago, Illinois.

Parts of each of the achievement test batteries mentioned above may be purchased separately for testing arithmetic; the California and Iowa at any level, and the Metropolitan and Stanford for grades five - eight.

INTELLIGENCE TESTS, GROUP

California Short-form Test of Mental Maturity, (Kindergarten - grade three), California Test Bureau, Los Angeles, California.

California Test of Mental Maturity, (Kindergarten - adult), California Test Bureau, Los Angeles, California.

Detroit Advanced First Grade Intelligence Test, (late first grade - early second grade), World Book Company, Yonkers-on-Hudson, New York.

Detroit Beginning First Grade Intelligence Test,

(Beginning first grade), World Book Company, Yonkers-on-Hudson, New York.

Detroit Primary Intelligence Test, (Grades two - four), Public School Publishing Company, Bloomington, Illinois.

Kuhlman-Anderson Intelligence Test, (Grade one - adult), Educational Test Bureau, Minneapolis, Minnesota.

Pintner-Cunningham Primary Mental Test, (Kindergarten - grade two), World Book Company, Yonkers-on-Hudson, New York.

Pintner-Durost Elementary Test, (Grades two - three), World Book Company, Yonkers-on-Hudson, New York.

Pintner Intelligence Test, (Grades four - eight), Teachers College Bureau of Publications, Columbia University, New York, New York.

INTELLIGENCE TESTS, INDIVIDUAL

Arthur Performance Scale, (Ages five - sixteen), C. H. Stoelting Company, Chicago, Illinois.

Kuhlman Revision of the Binet-Simon Scale, (Ages three months - fifteen years), Warwick and York, Incorporated, Baltimore, Maryland.

Pintner-Patterson Performance Scale, (Ages four - sixteen), C. H. Stoelting Company, Chicago, Illinois.

Porteus Maze Test, (Ages three - fourteen), C. H. Stoelting Company, Chicago, Illinois.

Stanford-Binet Scale (New), (Ages two - adult), Houghton-Mifflin Company, Boston, Massachusetts.

Wechsler-Bellevue Intelligence Scale, (Ages five - adult), The Williams and Wilkins Company, Baltimore, Maryland.

PERSONALITY TESTS

California Test of Personality, (Group; kindergarten - adult), California Test Bureau, Los Angeles, California.

READING READINESS TESTS

Gates Reading Readiness Test, (Kindergarten - first

grade), Teachers College Bureau of Publications, Columbia University, New York, New York.

Lee-Clark Reading Readiness Test, (Kindergarten - first grade), California Test Bureau, Los Angeles, California.

Metropolitan Readiness Test, (Kindergarten - first grade), World Book Company, Yonkers-on-Hudson, New York.

Monroe Reading Aptitude Test, (Beginning First Grade), Houghton-Mifflin Company, Boston, Massachusetts.

Stone-Grover Classification Tests For Beginners in Reading, (Beginning first grade), Webster Publishing Company, St. Louis, Missouri.

READING TESTS, GROUP

Durrell-Sullivan Reading Capacity Tests, (Grades two - six), World Book Company, Yonkers-on-Hudson, New York.

Gates Silent Reading Tests, (Grades one - eight), Teachers College Bureau of Publications, Columbia University, New York, New York.

Parts of each of the achievement test batteries listed above may be purchased separately for testing reading; the California and Iowa at any level, and the Metropolitan and Stanford for grades five - eight.

READING TESTS, INDIVIDUAL

Betts Ready to Read Battery, (Pre-school through college), The Psychological Corporation, New York, New York.

Gray's Standardized Oral Reading Check Tests, (Grades one - eight), Public School Publishing Company, Bloomington, Illinois.

SPELLING TESTS AND LISTS

Buckingham Extension of the Ayres Spelling Scale (Spelling list; grades two - nine), Public School Publishing Company, Bloomington, Illinois.

Iowa Spelling Scale, (Spelling list; grades two - eight),

Public School Publishing Company, Bloomington, Illinois.

Morrison-McCall Spelling Scale, (Group spelling tests; grades two - eight), World Book Company, Yonkers-on-Hudson, New York.

Parts of the Metropolitan and Stanford Achievement Test Batteries may be purchased separately for testing spelling at grade levels five - eight.